ON THE THEORETICAL STATUS OF "THE BASELINE" IN MODELLING INTONATION*

D. ROBERT LADD
University of Edinburgh

The "baseline", as an aspect of fundamental frequency (F_0) declination, is commonly defined as a straight (often logarithmic) line fitted to the valleys in the F_0 contour. In a recent study, in line with a good deal of past work on declination, Terken (1991) assumed that the baseline so defined would represent a reference value for normalizing the perceived prominence of pitch accents. The results of the study did not bear out this assumption, but the assumption was not called into question, and instead certain paradoxical conclusions were suggested. In this paper it is shown that the paradoxical conclusions are unnecessary if the reference value is not equated with the traditional baseline, but is assumed to be more abstract, i.e., less overtly manifested in the actual contour. This has already been proposed in a number of recent models of intonational phonology.

Key words: intonation, declination, prominence

INTRODUCTION

The term *declination* was coined a quarter of a century ago by Cohen and 't Hart (1967) as part of the development of a quantitative intonational model for speech synthesis at the Institute for Perception Research (IPO). The IPO researchers were not the first to notice the phenomenon (e.g., Pike, 1945, had already commented on the tendency of pitch to "drift" down gradually during the course of an utterance), but in many ways they set the agenda for experimental investigation of global trends in pitch contours. In particular, they seem to have been the first to focus on *topline* and *baseline* – lines fitted to peaks or valleys in utterance contours. A considerable number of subsequent studies on several languages have been devoted to investigating the properties of such lines (e.g., Maeda, 1976; Cooper and Sorensen, 1981; 't Hart, 1979; for a more general review of the IPO model in particular see 't Hart, Collier, and Cohen, 1990).

I will refer to this approach as the "overt decline" view. For researchers in this tradition, the essence of declination is a gradual decline in fundamental frequency (F_0) that is directly accessible to empirical observation and mathematically characterizable

---

* Carlos Gussenhoven, Alex Monaghan, Janet Pierrehumbert, Bruno Repp, and Jacques Terken made comments on an earlier draft of this paper which led to substantial improvements in the exposition.

Send correspondence to: Dr. D.R. Ladd, Dept. of Linguistics, University of Edinburgh, Adam Ferguson Building, Edinburgh EH8 9LL, UK.

as the slope of a regression line fitted to selected points in contours. Unfortunately for this view, empirical observation has not led to quantitative understanding. While there is a standard declination formula used for speech synthesis at IPO ('t Hart, 1979), it has not been widely adopted elsewhere, and there remain many inconsistencies and contradictions in the empirical literature. Summing up numerous studies, Vaissière (1983), herself a proponent of the overt-decline approach, remarks that "declination is a general tendency easily detected from a visual inspection of relatively long stretches of F_0 curve, but calculation of the exact rate of declination is a difficult task" (p. 56).

Since the early work of Fujisaki and his colleagues (e.g., Fujisaki and Sudo, 1971; Fujisaki and Hirose, 1982) there has existed an alternative to the overt-decline view, which I will refer to as the "implicit decline" approach. In this view what declines in declination is not necessarily F_0 itself, but rather an abstract backdrop against which F_0 is interpreted linguistically. For example, it is known that an accent peak of n Hz late in an utterance may be as "prominent" linguistically as a peak of $n + m$ Hz early in the same utterance (e.g., Pierrehumbert, 1979). The basis of the equivalence, rather than the slope of a regression line, is the focus of interest in the implicit-decline view. For example, in Fujisaki's model, contours are generated by adding the abstract specifications of an "accentual component" to the gradually declining value of the "phrase component", and it is the function that generates the phrase component that for Fujisaki is the true mathematical expression of declination. The overt baseline, in Fujisaki's view, is simply an epiphenomenon, which it is not only impossible but also theoretically incoherent to try to characterize quantitatively.[1]

More recently, the implicit-decline view has been adopted by Pierrehumbert and her colleagues (Pierrehumbert, 1979; Pierrehumbert, 1980; Liberman and Pierrehumbert, 1984; Pierrehumbert and Beckman, 1988; *cf.* also Ladd, 1984; Ladd, 1990; Gussenhoven and Rietveld, 1988). In much of this work, the emphasis is on determining and quantifying the linguistic equivalence of objectively different F_0 events at different points in a contour. Pierrehumbert's (1980) early model, for example, expressed the size of pitch excursions in "baseline units", whose size at any given point in an utterance was determined by the current level of an abstract declining "baseline". (Pierrehumbert's original model has been superseded in subsequent work, but for expository purposes it provides a simpler picture of the difference between the two approaches.) Like the IPO baseline, Pierrehumbert's baseline was a gradually declining straight line, and like the IPO researchers, Pierrehumbert hoped to relate baseline declination to a gradual decline in subglottal pressure or some similar physiological effect. However, for Pierrehumbert, the empirical determination of the baseline is not a matter of fitting a line to points in a contour, but rather a matter of finding the line that provides the *best expression of the linguistic equivalence of local F_0 features.* In Pierrehumbert's metaphor, declination is a kind of "tilting of the graph paper" on which F_0 contours are plotted;

---

1 Much the same point has been made for different reasons by Lieberman, Katz, Jongman, Zimmermann, and Miller (1985).

what we want to discover empirically is the amount by which the graph paper is tilted, not the downward trend of the plotted contour itself.[2]

The plausibility of the implicit decline view can perhaps be better appreciated with an analogy to music. In choral singing there is a tendency to go flat, i.e., a tendency for the pitch to decline gradually during stretches where the choir is unaccompanied by any instrument. It is instructive to consider what is meant by "going flat". It is not (except in the most egregious cases) a matter of an actual decline during the singing of any individual note. Nor is it a matter of a melody line that declines: "Going down" and "going flat" are conceptually clearly distinct, and in fact it is perfectly possible to go flat while singing a melody line that only rises. Rather, what declines when a choir goes flat is something like a frame of reference for interpreting pitches. For example, if a choir goes flat while singing a piece that starts and ends on the same note, then objectively they will be singing a lower pitch at the end than at the beginning. But only the most musically sensitive members of the audience are likely to notice; for the rest (as for the singers themselves) the pitch at the end and the pitch at the beginning count as the same.

If we had a graphic display of the melody actually sung by an inexpert choir, we could fit a gradually declining line to the display to characterize the rate at which they went flat. But it would obviously be essential for such a line to be fitted to points in the display that *count as the same note*. For example, it would not be meaningful to fit a line to the "valleys" in a melody line like the following (the refrain to the Christmas carol known in North America as "Angels We Have Heard On High"):

The decline in a line fitted to these valleys would not directly tell us anything about how badly the singers went flat, because the line is *supposed* to decline; the melody goes downward, but not only because of the decline in the singers' frame of reference.

For proponents of the implicit-decline approach to speech declination, the problem with the overt baseline fitted to valleys in intonation contours is much the same as that

[2] This aspect of declination has not gone unnoticed in the overt-decline view. Fairly early in the development of their model, the IPO researchers explicitly pointed out that the prominence of pitch accents in some way reflects their height relative to the baseline (e.g., Cohen, Collier, and 't Hart, 1982). Nevertheless, at least for ordinary declarative contours they have continued to define declination solely in terms of the actual F_0 at contour valleys. In this connection it should also be noted that the implicit-decline view is fairly neutral about the production mechanisms that underlie declination. The phrase component in Fujisaki's model is intended to be physiologically realistic, but its interpretation as an account of local prominence and pitch range does not depend on how accurately it models speech production.

with a line fitted to the valleys in the "Gloria" refrain. Such valleys may be easily identifiable points on a pitch trace, but there is no particular reason to believe that they are linguistically equivalent, and therefore no reason to believe that the overt baseline captures the essence of declination. It is just as likely that the overt baseline reflects aspects of both "going flat" and "going down", in indeterminate proportions. It is, of course, genuinely difficult to determine what counts as linguistically equivalent in a pitch contour, especially in languages without lexical tone. But that does not make it useful to cling to the overt baseline as a characterization of declination, just because it happens to be easy to define.

Against this background, a recent study by Terken (1991) is a potentially valuable addition to the experimental literature. Terken's stated aim was to explore the interaction of declination and local pitch range, specifically to find an appropriate metric for expressing the perceived prominence of pitch accents. By allowing subjects to set the prominence of one accent in a pitch contour equal to that of another, he obtained empirical data about the linguistic equivalence of objectively different F_0 events. Unfortunately, however, as an adherent of what I am calling the "overt-decline" approach, Terken simply assumed that the declining frame of reference for linguistic equivalence is directly signalled by the overt baseline. He tried to make sense of his data by normalizing accentual prominence to the baseline, and was unable to do so. Consequently, from Terken's point of view, his study was rather inconclusive; some of the results strike him as paradoxical, and in the discussion he airs a number of problems for which there are no obvious solutions.

My purpose in this paper is to argue that in fact Terken's experiments were a well designed and carefully executed test of the central prediction implicit in the overt-decline view of declination, and that this prediction fails. Specifically, I contend that Terken's experiment tested the hypothesis that the overt baseline does provide an adequate expression of the changing frame of reference for the linguistic equivalence of F_0 events, and that his results unambiguously contradict this hypothesis. Once we abandon the assumption that the overt baseline has any special role to play in characterizing declination, Terken's results lose their paradoxical nature. Moreover, we can go on to use his data to add to our understanding of the real essence of declination.

TERKEN'S PROMINENCE ADJUSTMENT EXPERIMENT

An overview of Terken's assumptions and results

Terken's procedures and basic results can be summarized briefly. In a two-accent synthetic utterance *ma–MA–ma–ma–ma–MA–ma*, Terken systematically varied the F_0 peak of the first accent (P1), and asked his subjects to manipulate the F_0 peak of the second accent (P2) so as to make its degree of prominence perceptually equal to that of P1. (There was another experimental task – setting the perceived *pitch* of P2 equal to P1 – which does not concern us here.) There were two experimental conditions, one in which the synthetic contours had a "flat baseline" and one in which they had a "declining baseline". That is, in one condition all the unaccented syllables of all contours had a fixed F_0 of 75 Hz, and in the other the F_0 of the unaccented syllables gradually declined through the utterance to reach a final F_0 of 75 Hz. In what follows the two conditions are referred to as the FBL (flat baseline) and DBL (declining baseline) conditions.

Two aspects of the DBL condition should be clarified from the outset. First, as in most work on F_0 done at IPO, pitch was expressed in semitones (st), so that the "declining baseline" was a straight line when plotted logarithmically. Second, as will be explained further below, the interval between the baseline and the various values of P1 was kept constant at six st. This means that the starting point of the baseline (and also its slope, because the endpoint was always 75 Hz) *increased* with the increasing values of P1. This is shown in Figure 1.

In general, consistent with related studies such as Pierrehumbert (1979) or Gussenhoven and Rietveld (1988), Terken found that his subjects set the F_0 of P2 lower than P1. The size of this effect – i.e., the difference between P1 and P2 – increased as the F_0 of P1 increased. The effect was also consistently greater in the DBL condition than in the FBL condition. These results are summarized in Figure 2. Before considering these results further, however, it will be useful to look first at how Terken presents the theoretical background to his study.

Terken sees two approaches to the problem of characterizing the prominence of accents. He designates these the MAX and CHANGE models; the former, he says, bases its notion of prominence on the actual F_0 of the accent peak, while the latter is based on the size of the F_0 excursion that manifests the accent. It is clear that neither of these models is adequate in its simplest version; interspeaker differences of range – to say nothing of any other complications – are enough to make either model impossible to apply in practice. The question is therefore how to modify the MAX and CHANGE models in order to make them workable.

Terken suggests that in a "more sophisticated" version of the CHANGE model one might calculate the size of the pitch excursion *with respect to the overt baseline of the contour.* The accent peak can be projected onto the baseline and the distance computed in an appropriate unit (e.g., semitones). These distances then become the expression of the prominence of the accent. He contrasts this sophisticated CHANGE model with a sophisticated MAX model based on Liberman and Pierrehumbert (1984), which may be regarded as a development of the Pierrehumbert (1980) model summarized above. In such a model, peak F_0 is mathematically transformed relative to an abstract reference line whose level depends in part on the overall pitch range of the utterance. These transformed peak values can then be taken as the expression of the prominence of any given accent.

In my view, by modifying the MAX and CHANGE models in this way Terken has eliminated any real distinction. I contend that the sophisticated version of MAX is conceptually equivalent to the sophisticated version of CHANGE. First, both models take it for granted that peak F_0 must be evaluated relative to the actual pitch range in which the contour is spoken. Second, both make use of reference values in the lower part of the speaking range,[3] relative to which F_0 peaks are normalized. And third, both

[3] It should perhaps be emphasized that the Liberman-Pierrehumbert model sharply distinguishes the "baseline", which is a fixed speaker-specific value for the bottom of the range, from the "reference line", which is well above the bottom of the range and increases as overall pitch range increases.

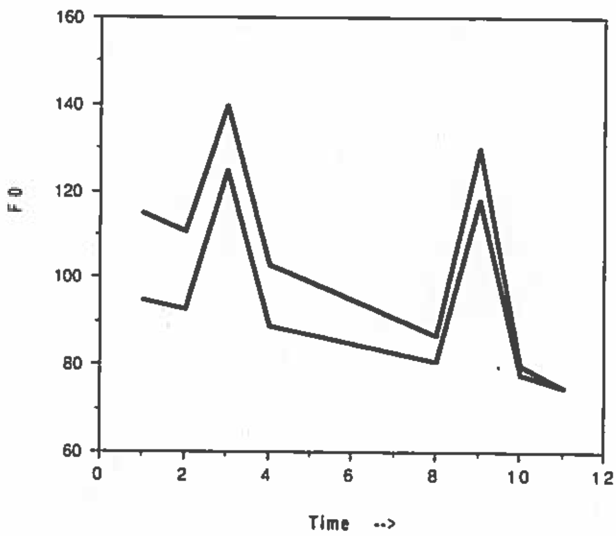

Fig. 1. Idealized representations of two different contours from Terken's DBL condition, showing how the starting point and slope of the overt baseline increase with the height of P1.

models assume that the reference value is subject to change – in the case at hand, subject to declination – so that it makes sense to talk about "local" pitch range, the pitch range available at any given point in an utterance. There are differences of detail, of course: Terken's sophisticated CHANGE model is based on a simple arithmetic difference between accent peak and overt baseline, whereas the sophisticated MAX model involves more complex mathematical manipulations of the F_0 space in which pitch movements are realized. But the two models do not disagree fundamentally about how to normalize prominence.

Where they do differ is in the way they think about their reference value. Terken, following 't Hart *et al.* (1990), considers the reference value to be a simple empirical property of utterance contours – the baseline. *A priori*, the fact that one can often fit a straight line to the valleys of a log-transformed F_0 contour does not make that line any more perceptually relevant than any other line one might plausibly fit to the contour. Yet because the Liberman-Pierrehumbert reference line is patently an abstraction, and because the overt baseline seems to characterize declination in a straightforward empirical way, Terken is led to think that he is doing something fundamentally different from Liberman and Pierrehumbert: He is empirically measuring pitch excursions (CHANGE) while they are mathematically transforming peak heights (MAX).

In my view, the differences Terken focuses on are minor. The crux of the issue is that Liberman and Pierrehumbert assign no special status to the overt baseline; the fact that it often declines is one fact among many that will someday be accounted for

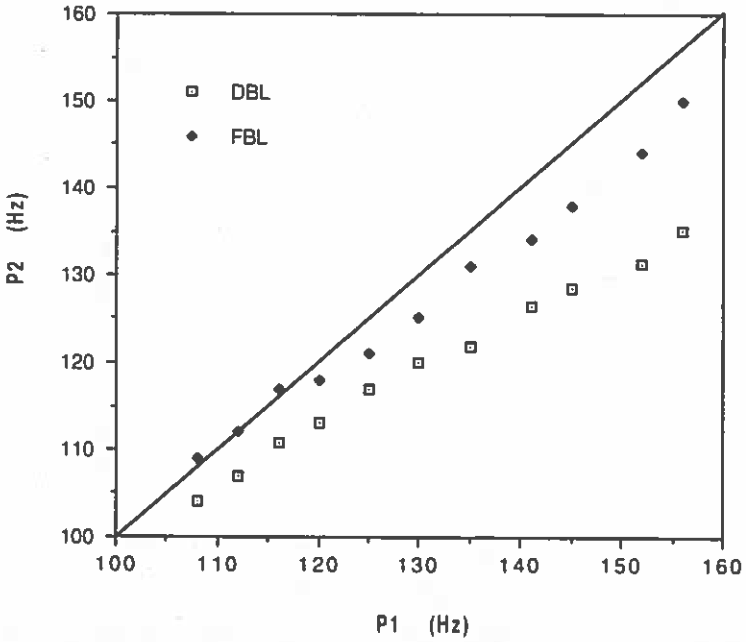

Fig. 2. Basic findings of the "prominence" task in Terken's study. The two sets of plotted points show the mean F_0 to which listeners set the peak of the second accent in an utterance (P2) in order to make it sound as prominent as a fixed first accent of a given F_0 (P1). DBL is the declining baseline condition and FBL the flat baseline condition. The straight line shows P2 = P1. This figure combines Terken's Figures 4 and 7.

in a comprehensive model of declination. But for Terken, the overt baseline is privileged. It is the central manifestation of declination, and related phenomena (like equivalent prominence of objectively different accents) must be accounted for within a framework of assumptions that is founded on the baseline's overt decline.

Terken's paradoxes

Consider now Terken's presentation of some of his central findings.

No equality of normalized prominence. In the DBL condition, Terken has taken care to construct his synthetic contours in such a way that P1 is always six st above the overt baseline. In his terms, D1 (the distance of P1 above the baseline) is always six st. (Actually it varies slightly for technical reasons, but this does not affect the basic point.) Since he assumes that prominence is normalized to the baseline, he expects that the experimentally determined values of P2, when measured with respect to the baseline, should display a similar constancy. In fact, however, the distance of P2 above the

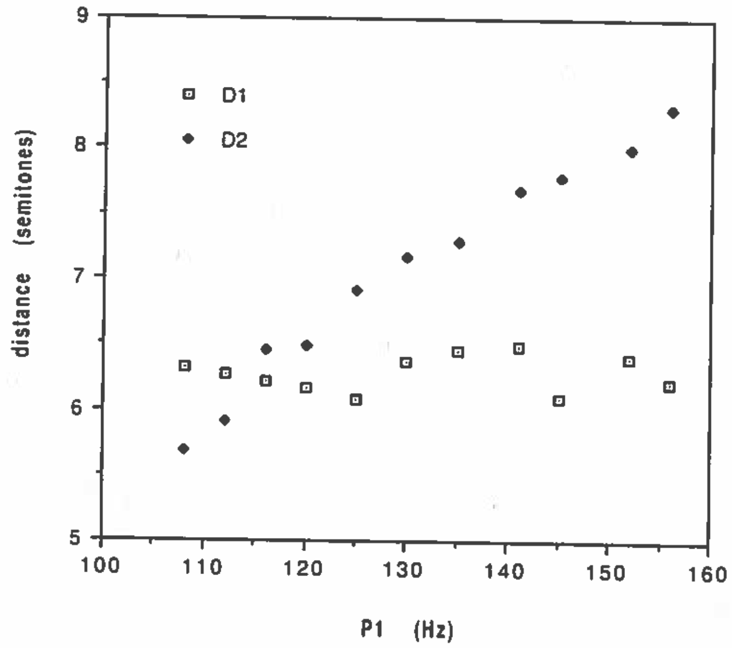

Fig. 3. Distance in semitones above the IPO baseline, for all values of both P1 (fixed) and P2 (set by subjects), in the DBL condition of Terken's study. D1 is the distance above the baseline of P1, which Terken held constant at about six st (the minor variations shown arose for technical reasons). D2 is the distance above the baseline of P2, which Terken expected to be constant as well. The gradually increasing value of D2 shown in the figure is the central issue dealt with in both Terken's original paper and here. Based on Terken's Table I.

baseline – D2 in Terken's terms – gradually increases as the absolute F_0 of P1 and P2 increases (*cf.* Figure 3). Terken finds it "puzzling" that the perceptually equal prominence of the two accents cannot be expressed in these terms, but never calls into question the use of the baseline as the reference value for normalizing peak height. Instead, he goes on to analyse his data with this assumption intact, and consequently finds himself reaching more and more problematic conclusions.

A linear perceptual scale? In Figure 3, the values of D1 and D2 are plotted in st, but the same inequality of D1 and D2 is observed whether D2 is measured in st, Hz, or the ERB units used by Hermes and van Gestel (1991). In fact, the divergence is by far the greatest with st, in the sense that the st measurements show no correlation between D1 and D2. Table 1 shows the constants in the regression functions that relate D1 and D2 in the three different units. Terken notes that both ERB and Hz display good

TABLE 1

Constants in the linear regression functions relating D1 to D2 for three different units of measurement for F_0, in Terken's original analysis of the DBL condition (from Terken's Table I)

	Unit of F_0 measurement		
	st	ERB	Hz
correlation	0.02	0.92	0.96
y-intercept	1.35	–0.95	–14.99
slope	0.91	1.96	1.39

correlations between D1 and D2, but that the slope of the regression line for Hz is closer to 1.00 (i.e., closer to expressing equality) than the one for ERB.

Already puzzled by the failure of D1 and D2 to reflect the perceptual equality of P1 and P2, Terken finds it even more puzzling that it is the linear Hz scale – which is presumably irrelevant psychoacoustically – that comes closest to making the accents look equal. As we shall see shortly, this puzzle is purely an artifact.

A fundamental effect of "declination"? Comparison of the DBL and FBL conditions leads Terken to perceive yet another paradox. In FBL, D2 and D1 both increase as the pitch range increases, but D2 does not increase as fast as D1. That is, the difference between D1 and D2 becomes greater as the pitch range gets wider; graphically, the value of D1–D2 increases as a function of P1. By contrast, as we just saw, in DBL the value of D2 increases as the pitch range increases, but not the value of D1, which Terken held constant. In DBL, that is, D1–D2 has an increasing *negative* value as a function of P1. This is shown in Figure 4.

Terken recognizes that this is a problem: "The CHANGE hypothesis would have to be formulated in opposite ways for utterances with and without baseline declination" (p. 1774). However, as a solution he proposes only that "expectations about the course of the topline are affected to a certain extent by the presence of audible baseline declination". He makes no attempt to quantify these "expectations", and once again does not call into question the validity of "distance above the [overt] baseline" as an appropriate measure of prominence.

A different point of departure

In my view, the paradoxes just outlined are almost entirely spurious. They arise because "distance above the baseline" is not (as Terken simply assumes) a straightforward empirical measure of pitch excursion, but rather a specific hypothesis about how the

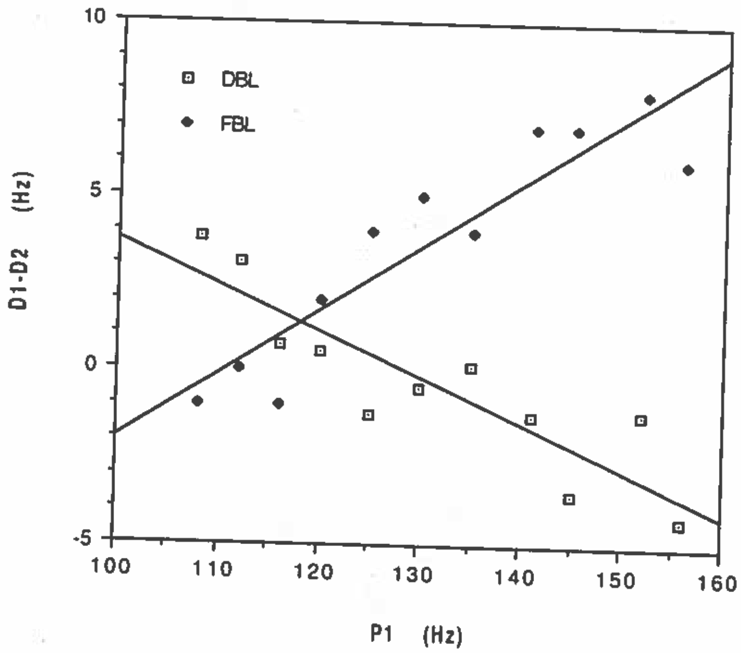

Fig. 4. Difference between D1 and D2 in Hz, for both the DBL condition and the FBL condition, in Terken's original data. (*Cf.* Terken's Table 1, his discussion p. 1774, and Figure 6 below.)

prominence of accent peaks is to be characterized. Like any hypothesis, this one can be falsified. Once this is accepted, the implications of Terken's experimental results become clear. We do not need to accept the puzzling conclusion that a linear frequency scale is the best choice for measuring the prominence of accents; we do not need to account for a fundamental divergence between the FBL and DBL conditions. We simply need to recognize that, in evaluating peak height, the baseline is *the wrong place to measure from.* There may well be a simple reference line that can serve the function of normalizing local pitch range, but it cannot be determined by inspection of the valleys in the F_0 contour. This point of view is developed in the next part of the paper.

A Reinterpretation of Terken's Results

A preliminary reanalysis

This section presents the results of a limited reanalysis of Terken's data. The reanalysis is "preliminary" in that its main purpose is to show that Terken's paradoxes are easily

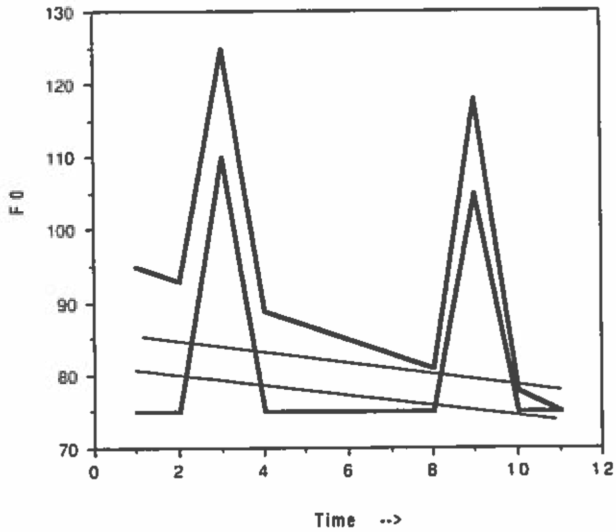

Fig. 5. The basis of the reanalysis of Terken's results. A reference line with a fixed slope of 3.4 st/second is used for all cases; this reference line is set to intersect the overt baseline at the point where P2 is projected onto the overt baseline, so that D2 is the same in both the reanalysis and in Terken's original analysis, and only D1 differs. The higher reference line is for the DBL contour, the lower reference line for the FBL contour.

resolved. It is "limited" in the sense that it preserves most of Terken's basic assumptions. First, it assumes that prominence is normalized relative to a reference value low in the speaking range. Second, it also follows Terken in expressing the normalization simply as an arithmetic difference between the accent peak and its reference value (rather than the more complicated normalizations proposed by Liberman and Pierrehumbert, 1984, or Ladd, 1990). In fact the only modification made to Terken's procedures is to assume a reference value other than the overt baseline. Even with this very modest change, the paradoxes set forth above disappear.

In the reanalysis, the reference value for *all* the test utterances in *both* the FBL and DBL conditions is an abstract line declining at the standard declination rate specified by the IPO speech synthesis model. Since the test utterance is 1680 msec long, the rate of declination is 3.4 st/second, which amounts to 3.11 st decline in the 900 msec interval between P1 and P2. In order to make the reanalysis as comparable as possible with Terken's original analysis, this baseline has been "anchored" at the baseline value assumed by Terken for P2. (In what follows, this point – the point on the overt baseline onto which P2 is projected – will be referred to as R2.) This means that D2 (i.e., P2 – R2) is identical in both the reanalysis and Terken's original analysis; only D1 is different.

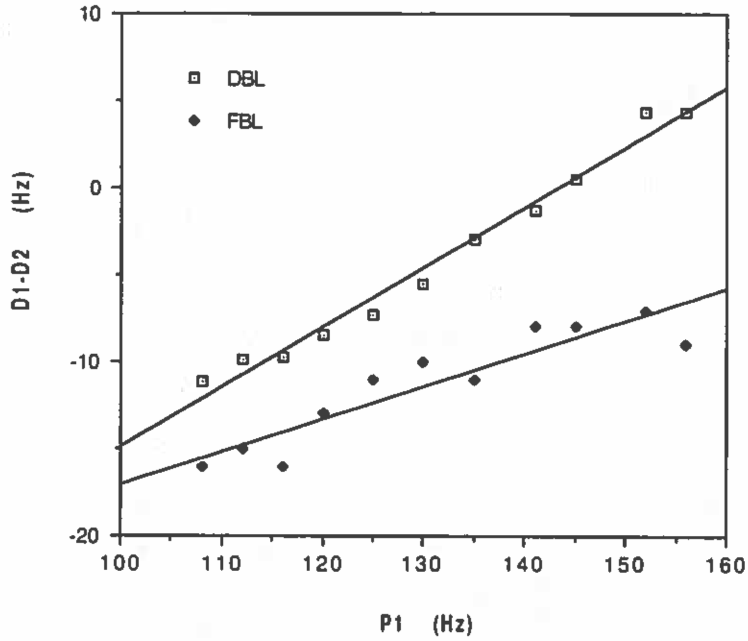

Fig. 6. Difference between D1 and D2 in Hz, for both the DBL condition and the FBL condition, in the reanalysis of Terken's data. (*Cf.* Figure 4 above.)

The basis of the reanalysis is shown schematically in Figure 5.[4]

The results of the reanalysis show clearly that Terken's paradoxes are spurious. First, the apparent divergence of the flat baseline and declining baseline conditions disappears. In both conditions, the difference between D1 and D2 gradually increases as P1 increases. This is shown in Figure 6 (which should be compared to Figure 4). While this is a step

---

[4] I wish to emphasize that I am not proposing this as the most plausible set of assumptions about the nature of declination. The main purpose of the reanalysis is to show that, given a set of assumptions that are not wildly implausible but still different from Terken's overt-decline assumptions, the supposedly paradoxical aspects of his results disappear. Note in particular that "anchoring" the declination line to the point R2 was done primarily for convenience and accuracy, since it made it simple to calculate R1 based only on the information provided in Terken's published paper. With a good deal more computation and a certain amount of possibly unjustified extrapolation from the details given by Terken, it would have been possible to estimate the location of a declination line anchored to the contour end point of 75 Hz, which would certainly have been more in line with a number of models. However, the quantitative differences between the reanalysis presented here and one based on the more conventional declination line would have been minimal. Consequently this reanalysis may be taken more generally as a fair test of standard IPO assumptions about declination and prominence.

TABLE 2

Constants in the linear regression functions relating D1 to D2 for three different units of measurement for F_0, in the reanalysis (DBL condition only)

	Unit of F_0 measurement		
	st	ERB	Hz
correlation	0.987	0.991	0.989
y-intercept	4.02	0.57	20.01
slope	0.561	0.574	0.562

in the right direction, the fact that D1 and D2 are not equal means that the reanalysis still fails to express the linguistic equivalence of the two accents. This point is taken up in the next section.

Second, the divergence among Hz, ERB, and st as units for expressing D1 and D2 is much less stark than in Terken's original analysis. For all three units, the correlation of D1 and D2 is nearly perfect, and the slopes of the regression lines differ little. This is shown in Table 2 (which should be compared with Table 1). A little reflection on these results makes it clear why Terken found such large divergences among the three units of measurement: He set the baseline at precisely the point where the prominence of P1 is made to appear unchanging in one unit but not in the others. The absence of correlation between D1 and D2 in st in Table 1 is meaningless, a mathematically inevitable consequence of normalizing to that particular baseline. With a different baseline, the very high correlation reappears.

Finally, the reanalysis provides a substantive finding: Irrespective of the unit of measurement in which D1 and D2 are expressed, D1 and D2 both increase with P1 and P2. That is, *both accents increase in prominence as the overall pitch range increases.* Pretheoretically this makes a certain amount of sense, and could be made to follow from certain assumptions about the nature of the reference value for normalizing prominence. But in effect, Terken's attempt to build equal prominence into P1 in the DBL condition (by making P1 always six st above the baseline) can be seen as a cleverly designed test of the *opposite* prediction. Terken assumed (which is simply a strong form of predicting) that if he raised the level of the unstressed syllables in a particular way he would thereby directly raise the level of the reference value, and would thus *hold the prominence of P1 constant as overall pitch range increased.* This assumption – or prediction – is what his subjects' behavior conspicuously fails to support.

Learning from Terken's results

The reanalysis thus makes clear that Terken's paradoxical conclusions are unwarranted, and that his central assumption is false. It also shows that a hypothetical declination line with IPO-standard slope, because it fails to express the perceptual equality of the accent peaks, does not capture the basis of the subjects' behavior much better than the overt baseline. However, negative conclusions are not the net result of Terken's study. On the contrary, given some set of assumptions about how to normalize prominence, we can actually use Terken's data to establish *empirically* how steep the declination of the subjects' reference value must have been. That is, since we know that the two accent peaks were perceived to be equally prominent, we can infer the reference value for any P1 on the basis of the reference value and the normalized prominence of the corresponding P2.

Suppose, for example, we retain Terken's simple arithmetic normalization; for simplicity, suppose we also retain the reference value for P2 to which we "anchored" the reanalysis in the previous section. That is, suppose we assume that the prominence values for the various values of P2 in both the DBL and FBL conditions are the values of D2 that we have used throughout. It is trivial to work backwards from these values and to determine, for any given P1, where the reference value – call it R1 – *must be*. These values of R1 are shown in Table 3.[5]

Naturally, the values in Table 3 are only as good as the assumptions about normalization on which they are based. That is, *if* there is indeed a declining reference line low in the range, and *if* the true measure of prominence is simply the arithmetic difference in ERB units between the peak and the declining reference line, then Table 3 gives us a good idea of the slope and approximate location of the reference line. If those assumptions are not valid, then the details of Table 3 are probably irrelevant. But the approach underlying Table 3 seems useful regardless of our starting assumptions: Given a set of assumptions about how to normalize prominence, we can use Terken's data to work backwards from the reference value for one accent to the reference value for the other. Thus Terken's data can be used to help refine models of pitch range and prominence normalization based on a variety of assumptions.

Terken's quantitative findings lend themselves less easily to identifying sets of assumptions about normalization that are *a priori* plausible – that is, deciding which sets of starting assumptions it actually makes sense to explore. For example, even if the assumptions about the nature of normalization in Table 3 are more or less valid, the results give us little basis for generalizing to other cases. We have determined where R1 must be, but we know nothing more about what relates R1 to R2. For example, we might continue to assume that R1 and R2 lie along a straight declining line like the one the IPO model has always used – i.e., like the overt baseline, but not fitted to the valleys

---

[5] Readers who are disturbed by the fact that R1 rises with P1 even in the FBL condition should keep in mind that this is not some perverse theoretical consequence of the implicit-decline view, but an *empirical finding*. That is, even when the overt baseline is flat, there is still declination in the sense that subjects consistently set P2 lower than P1 to achieve equal prominence. The implicit decline view attempts to account for this finding by not assuming that the overt baseline tells us everything we need to know about the rate of declination.

TABLE 3

Empirically determined reference values (R1) for P1, in both the DBL and FBL conditions. These values are based on the assumption that D1 and D2 are equal in ERB (reflecting the perceived equality of prominence); they are derived by subtracting Terken's observed D2 from P1 in ERB, and converting the result to Hz

		DBL condition		FBL condition	
P1 (Hz)	P1 (ERB)	observed D2 (ERB)	estimated R1 (Hz)	observed D1 (ERB)	estimated R1 (Hz)
108	3.64	0.83	78	0.96	74
112	3.75	0.87	81	1.04	75
116	3.85	0.96	81	1.17	74
120	3.96	0.98	84	1.20	77
125	4.08	1.05	86	1.27	78
130	4.21	1.11	88	1.37	79
135	4.33	1.14	92	1.52	78
141	4.47	1.22	94	1.59	81
145	4.57	1.25	96	1.69	81
152	4.73	1.29	101	1.83	82
156	4.82	1.36	101	1.97	80

of the contour, and with a slope that varies as a function of pitch range. Or we might assume that R1 and R2 lie along an asymptotically decaying line like Fujisaki's phrase component. Or perhaps there is no simple time function describing declination at all; perhaps, as suggested by Pierrehumbert and others, the lower F_0 of P2 reflects a distinct "final lowering" of the reference value (*cf.*, e.g., Liberman and Pierrehumbert, 1984). The figures in Table 3 do not, on their own, give us any basis for deciding among these approaches.

But even in guiding the choice of starting assumptions Terken's results are of some use. For example, as we saw in the previous section, it seems clear that overall increases in peak height (relative to the absolute bottom of the range, not relative to any overt baseline) yield overall increases in prominence. The reference value (if there is such a thing) must therefore go up less rapidly than the height of peaks. This finding can confidently be used in building a general model of pitch range. Similarly, on the basis of the difference between DBL and FBL, it also seems well justified to conclude that the reference value increases with the height of the overt baseline.[6] To put it another way,

[6] This is seen clearly in Table 3, but the same relationship between overt baseline and reference value would obtain given a wide variety of starting assumptions, not just the specific ones underlying Table 3.

a given peak level is less prominent when there is a high overt baseline (as in DBL) than when there is a low one (as in FBL) – presumably because the peak is less distant from the reference value. This conclusion, if well-founded, favors a set of assumptions like those adopted by Liberman and Pierrehumbert (1984), or by Ladd (1990). Specifically, one could readily incorporate this conclusion into a model whose reference value lies *between* the valleys and the peaks – which is where both Liberman and Pierrehumbert's reference line and the "zero line" in the middle of Ladd's "tonal space" are located. But it is difficult to accommodate it in any model that takes the valleys alone – the overt baseline – as the basis for normalizing local pitch range.

CONCLUSION

I have argued in this paper that Terken's prominence adjustment study, far from providing inconclusive and paradoxical results, has given us a clear test of the assumption that the baseline – in the sense of a line fitted to the valleys in an F_0 contour – is the appropriate reference value for comparing the relative prominence of pitch accents. The result is unmistakably negative: The overt baseline does not work in the way predicted, and some more abstract "implicit decline" approach to describing the effects of declination on local pitch range is required.

The principal lesson to be drawn from Terken's experiment, I have suggested, is that the reference value for normalizing local pitch range does not lie along a line fitted to either the peaks or the valleys of the contour, but is an abstract construct somewhere between the valleys and the peaks. As I noted, this is entirely consistent with the "implicit-decline" models of e.g., Liberman and Pierrehumbert (1984) or Ladd (1990). Terken (1993) accepts this general point in his reply to this paper; as he notes, both the peaks and the valleys are relevant to the listener in inferring the location of the reference value.

More generally, I have tried to show that we need to strip the overt baseline of its special status in F_0 research. We need to acknowledge that the key to normalizing prominence is some sort of abstract reference value in a comprehensive model of pitch range. The baseline drawn through F_0 valleys, however salient its decline may be, is only one manifestation of how pitch range evolves during the act of speaking.

(Received September 28, 1992; accepted March 31, 1993)

REFERENCES

COHEN, A., COLLIER, R., and 'T HART, J. (1982). Declination: construct or intrinsic feature of speech pitch? *Phonetica*, **39**, 254–273.

COHEN, A., and 'T HART, J. (1967). On the anatomy of intonation. *Lingua*, **19**, 177–192.

COOPER, W., and SORENSEN, J. (1981). *Fundamental Frequency in Sentence Production*. Heidelberg, Germany: Springer-Verlag.

FUJISAKI, H., and HIROSE, K. (1982). Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In Preprints of Papers, Working Group on Intonation, 13th International Congress of Linguists, Tokyo, pp. 57–70.

FUJISAKI, H., and SUDO, H. (1971). A generative model for the prosody of connected speech in Japanese. *Annual Report of the Engineering Research Institute* (University of Tokyo), **30**, 75–80.

GUSSENHOVEN, C., and RIETVELD, T. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, **16**, 355–369.

HERMES, D., and VAN GESTEL, J. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, **90**, 97–102.

LADD, D.R. (1984). Declination: A review and some hypotheses. *Phonology Yearbook, 1*, 53–74.

LADD, D.R. (1990). Metrical representation of pitch register. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology, I* (pp. 35–57). Cambridge, U.K.: Cambridge University Press.

LIBERMAN, M., and PIERREHUMBERT, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.), *Language Sound Structure* (pp. 157–233). Cambridge, MA: MIT Press.

LIEBERMAN, P., KATZ, W., JONGMAN, A., ZIMMERMANN, R., and MILLER, M. (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *Journal of the Acoustical Society of America*, **77**, 649–657.

MAEDA, S. (1976). *A Characterization of American English Intonation.* PhD. dissertation, MIT.

PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363–369.

PIERREHUMBERT, J. (1980). *The Phonology and Phonetics of English Intonation.* PhD. dissertation, MIT.

PIERREHUMBERT, J., and BECKMAN, M. (1988). *Japanese Tone Structure.* Cambridge MA: MIT Press.

PIKE, K.L. (1945). *The Intonation of American English.* Ann Arbor, MI: University of Michigan Press.

TERKEN, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, **89**, 1768–1776.

TERKEN, J. (1993). Baselines revisited: Reply to Ladd, *Language and Speech*, **36**, 453–459.

'T HART, J. (1979). *Naar automatisch genereeren von toonhoogte-contouren voor tamelijk lange stukken spraak.* IPO Technical Report No. 353. Eindhoven: IPO.

'T HART, J., COLLIER, R., and COHEN, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach.* Cambridge, U.K.: Cambridge University Press.

VAISSIÈRE, J. (1983) Language-independent prosodic features. In A. Cutler and D.R. Ladd (eds.), *Prosody: Models and Measurements* (pp. 53–66). Heidelberg, Germany: Springer-Verlag.